HD

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/695,077	10/25/2000	PHILIP NEIL GARNER	1263.1881	1606

5514 7590 03/26/2007
FITZPATRICK CELLA HARPER & SCINTO
30 ROCKEFELLER PLAZA
NEW YORK, NY 10112

EXAMINER	
LERNER, MARTIN	

ART UNIT	PAPER NUMBER
2626	

SHORTENED STATUTORY PERIOD OF RESPONSE	MAIL DATE	DELIVERY MODE
3 MONTHS	03/26/2007	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

If NO period for reply is specified above, the maximum statutory period will apply and will expire 6 MONTHS from the mailing date of this communication.

Office Action Summary	Application No. 09/695,077	Applicant(s) GARNER ET AL.	
	Examiner Martin Lerner	Art Unit 2626	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on 28 December 2005.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) 1 to 95 and 97 to 102 is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) 1-5,20-25,30-41,46,48-52,67-73,77-88,90,93 and 97-102 is/are rejected.

7)☒ Claim(s) 6-19,26-29,42-45,47,53-66,74-76,89,91,92,94 and 95 is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☒ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☒ All b)☐ Some * c)☐ None of:

1.☒ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)

2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)

3) ☒ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date _____.

4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.

5) ☐ Notice of Informal Patent Application

6) ☐ Other: _____.

DETAILED ACTION

Claim Rejections - 35 USC § 103

1. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

2. Claims 1 to 5, 21 to 25, 30 to 32, 37 to 39, 48 to 52, 68 to 73, 77 to 79, 84 to 86, 90, 97, 99, and 101 are rejected under 35 U.S.C. 103(a) as being unpatentable over *Chou et al.* in view of *Goldberg et al. ('158)*.

Concerning independent claims 1, 48, 97, 99, and 101, *Chou et al.* discloses an apparatus, method, and instructions, comprising:

"a receiver operable to receive an input signal" – an input of an unknown speech string 18 (an utterance) of words is received from a microphone (column 4, lines 34 to 35: Figure 1);

"a recognition processor operable to compare said input signal with stored label models to generate a recognized sequence of labels in said input signal and confidence data representative of the confidence that the recognized sequence of labels is representative of said input signal" – recognition processor 10 receives the input, accesses the recognition database 12, scores the unknown speech string of words against the recognition models in the recognition database 12, and generates a

hypothesis string signal 20; verification processor 16 receives the hypothesis string, and generates a confidence measure signal 22 (column 4, lines 34 to 51: Figure 1);

"a similarity measure calculator operable to compare said recognized sequence of labels received from said recognition processor with a stored sequence of labels using a combination of i) predetermined confusion data which defines confusability between different labels, and ii) said confidence data received from the recognition processor and representative of the confidence that said received recognized sequence of labels is representative of the input signal, to provide a measure of the similarity between the recognized sequence of labels and the stored sequence of labels" – confidence score computation ("a similarity measure calculator") for a speech segment q relates a comparison between a word model score ("said confidence data") and scores computed with the anti-word model ("predetermined confusion data which defines confusability between different labels"); in Equation (2), $L(O_q;\Theta,l)$ is "the measure of similarity" calculated by the similarity measure calculator, $g_l(O_q) = log\ p(O_q|\theta_l^{(k)})$ is "the confidence data" for the keyword hypothesis $\{\theta_l^{(k)}\}$, and $G_l(O_q)$ is the "predetermined confusion data which defines confusability" for anti-keywords $\{\theta_l^{(a)}\}$ which handle confusibility among keywords (column 8, lines 33 to 55: Figure 2).

Concerning independent claims 1, 48, 97, 99, and 101, *Chou et al.* discloses scores computed with an anti-word model in Equation (2), where, arguably, $G_l(O_q)$ is the "predetermined confusion data which defines confusability" for anti-keywords $\{\theta_l^{(a)}\}$ which handle confusibility among keywords (column 8, lines 33 to 55: Figure 2). However, even assuming *arguendo* that "predetermined confusion data which defines

confusability between labels" is not clearly disclosed by *Chou et al.*, it is still well known to account for confusability between similarly sounding letters in speech recognition. Specifically, *Goldberg et al. ('158)* teaches a statistical option generator for speech recognition, where confusion sets are defined for groups of letters that have a certain probability of being confused with one another as represented by confusion matrices. (Column 3, Line 55 to Column 4, Line 10) Letters "A", "J", and "K" are grouped together in Confusion Set 1, letters "B", "C", "D", "E", "P", "T", and "V" are grouped together in Confusion Set 2, etc. (Column 8, Lines 14 to 25: Figure 4) An objective is to reduce the time for matching an input identifier and conserving computing power by eliminating option identifiers from a candidate set of reference identifiers. (Column 3, Lines 37 to 54) It would have been obvious to one having ordinary skill in the art to take into account confusion probabilities as taught by *Goldberg et al. ('158)* in a speech recognition system and method of *Chou et al.* for a purpose of reducing time for matching and conserving computing power.

Concerning claims 2 and 49, *Chou et al.* discloses the confidence measure is generated based upon data stored in verification database 16 (column 4, lines 34 to 51: Figure 1).

Concerning claims 3 and 50, *Chou et al.* discloses a word-based confidence score 34 (column 8, lines 40 to 55: Figure 2); each word is a "label" in a string of words being recognized.

Concerning claims 4 and 51, *Chou et al.* discloses string models are generated in an "N" best list ("a list of alternatives") by N-best string model generator 46 (column 6, line 45 to column 7, line 53: Figure 2; column 9, line 54 to column 10, line 14).

Concerning claims 5 and 52, *Chou et al.* discloses Viterbi alignment ("an aligner") of the input string, O, against the model sets for each given word string in the N-best string list (column 7, lines 8 to 15); average word-based confidence score processor 36 ("a combiner") performs mathematical averaging for each word segment signal of the hypothesis string to generate an average word-based confidence score signal ("said similarity measure") (column 5, lines 53 to 67: Figure 2).

Concerning claims 21 and 68, *Chou et al.* discloses Viterbi alignment ("an aligner") of the input string, O, against the models sets for each given word string in the N-best string list (column 7, lines 8 to 15); Viterbi alignment is "a dynamic programming technique".

Concerning claims 22 to 25, 69 to 72, and 90, *Chou et al.* discloses Viterbi alignment ("an aligner") of the input string, O, against the model sets for each given word string in the N-best string list (column 7, lines 8 to 15); implicitly, Viterbi alignment determines "progressively a plurality of possible alignments", generates scores for each given word in the N-best list, determines "an optimum alignment", and "combines the scores" for each word in the word string.

Concerning claims 30 to 32 and 77 to 79, *Chou et al.* discloses the input string is speech (column 4, lines 34 to 35: Figure 1), which is a time sequential audio signal of words.

Concerning claims 37, 38, 84, and 85, *Chou et al.* discloses confidence score computation for a speech segment q relates a comparison between a word model score ("said confidence data") and scores computed with the anti-word model ("said confusion data"); in Equation (2), $L(O_q;\Theta,l)$ is "the measure of similarity" calculated by the similarity measure calculator, $g_l(O_q) = \log p(O_q|\theta_l^{(k)})$ is "the confidence data" for the keyword hypothesis $\{\theta_l^{(k)}\}$, and $G_l(O_q)$ is "the confusion data" for anti-keywords $\{\theta_l^{(a)}\}$ which handle confusibility among keywords (column 8, lines 33 to 55: Figure 2).

Concerning claims 39 and 86, *Chou et al.* discloses an average confidence score based on upon the average of word-based confidence scores (column 5, lines 53 to 67); an average confidence score is a normalization from each of the word-based confidence scores.

Concerning claim 73, *Chou et al.* discloses each of the words ("labels") in the unknown speech string ("each of the labels in said recognized sequence of labels") is scored against recognition models ("stored sequences of labels") in the recognition database 12 (column 4, lines 34 to 51: Figure 1).

3. Claims 20, 35, 36, 46, 67, 82, 83, 93, 98, 100, and 102 are rejected under 35 U.S.C. 103(a) as being unpatentable over *Chou et al.* in view of *Goldberg et al. ('158)* as applied to claims 1, 5, 48, 52, 97, 99, and 101 above, and further in view of *Aref et al.*

Concerning claims 20 and 67, *Chou et al.* omits an aligner operable to identify deletions and insertions. However, *Aref et al.* teaches an analogous art speech recognition system for correcting misspelled words in a string of text. (Column 1, Lines

32 to 52) Specifically, *Aref et al.* discloses detecting recognition errors as models from insertion errors and deletion errors. (Column 3, Lines 36 to 60) It is suggested that there are advantages to speed the search process and reduce the size of the database by correcting misrecognized or misspelled words with the search technique of *Aref et al.* (Column 1, Lines 52 to 61) It would have been obvious to one having ordinary skill in the art to incorporate the insertion and deletion error technique of *Aref et al.* into the word-based confidence score method of *Chou et al.* for the purpose of correcting misrecognitions with a high speed search process and reduced database size.

Concerning claims 35, 36, 82, and 83, *Chou et al.* omits mis-typing probabilities and mis-spelling probabilities based upon sub-word units. However, *Aref et al.* teaches an analogous art speech recognition system for correcting misspelled words in a string of text. (Column 1, Lines 32 to 52) Specifically, *Aref et al.* discloses probabilities for letters being recognized incorrectly, where letters are sub-word units, to estimate a measure of similarity between two words. (Column 4, Lines 1 to 59) Recognition errors are based upon typing errors, e.g. "airnmail" is mistakenly inserted for the word "airmail". (Column 3, Lines 36 to 53) It is suggested that there are advantages to speed the search process and reduce the size of the database by correcting misrecognized or misspelled words with the search technique of *Aref et al.* (Column 1, Lines 52 to 61) It would have been obvious to one having ordinary skill in the art to utilize the mis-typing and mis-spelling technique of sub-word units taught by *Aref et al.* into the word-based confidence score method of *Chou et al.* for the purpose of

correcting misrecognitions with a high speed search process and reduced database size.

Concerning claims 46, 93, 98, 100, and 102, *Chou et al.* omits an application of speech recognition to querying a database and obtaining information from the database, although this is a well known application for speech recognition systems, generally. However, *Aref et al.* teaches an analogous art speech recognition system for searching a database for recognized text by querying keywords. (Column 2, Lines 41 to 50) It is suggested that there are advantages to speed the search process and reduce the size of the database by correcting misrecognized or misspelled words with the search technique of *Aref et al.* (Column 1, Lines 52 to 61) It would have been obvious to one having ordinary skill in the art to apply the word-based confidence score method of *Chou et al.* to a retrieval system from a database of automatically recognized text as taught by *Aref et al.* for the purpose of correcting misrecognitions with a high speed search process and reduced database size.

4. Claims 33, 34, 80, and 81 are rejected under 35 U.S.C. 103(a) as being unpatentable over *Chou et al.* in view of *Goldberg et al. ('158)* as applied to claims 1, 31, 32, 48, 79, and 80 above, and further in view of *Wheatley et al.*

Chou et al. discloses recognizing speech with word-based confidence scores, where the labels are words, but omits recognizing sub-word units and phonemes. However, it is a well known art recognized alternative in speech recognition to recognize phonemes, which are sub-word units, rather than words. *Wheatley et al.* teaches a

related apparatus and method for speech recognition, where speech is recognized with Hidden Markov Models representing phonetic units instead of words. (Column 7, Lines 14 to 37) It is suggested that there is an advantage of representing real world, unscripted conversations. (Column 2, Lines 28 to 39) It would have been obvious to utilize sub-word phonetic units for the speech recognition system of *Chou et al.* as suggested by *Wheatley et al.* for the purpose of better recognizing real world, unscripted conversations.

5. Claims 40, 41, 87, and 88 are rejected under 35 U.S.C. 103(a) as being unpatentable over *Chou et al.* in view of *Goldberg et al. ('158)* as applied to claims 1, 37, 39, 48, 84, and 86 above, and further in view of *Glenn et al.*

Concerning claims 40 and 87, *Chou et al.* discloses an average confidence score based on upon the average of word-based confidence scores (column 5, lines 53 to 67), where an average confidence score is a normalization from each of the word-based confidence scores, but does not expressly say that the normalization is obtained by dividing each similarity measure by a respective normalization score which varies in dependence upon the length of the corresponding stored sequence. However, *Glenn et al.* teaches a speech recognition system and method, where a pattern classifier 50 matches bits between an event coder output and candidate reference patterns from a reference pattern memory 60. Counter 55 represents the number of bits that match between the event coder output and the reference pattern, where the count of matching bits is denoted by C_k, and the maximum output from counter 55 is N_k, which represents

a perfect match. Score computer 56 calculates a value $S_k = M(C_k/N_k)$, which is the pattern classification score, and M is a constant equal to the maximum score possible. The ratio C_k/N_k "normalizes" the scores obtained to account for the consistency of the training patterns. (Column 6, Lines 31 to 68: Figure 1) Value S_k is a score, or "similarity measure", that is normalized "by dividing by a respective normalization score which varies in dependence upon the length of the correspond stored sequence" because N_k represents the length of the compared patterns, as the maximum number of matching bits is the length of the corresponding patterns, and score, S_k is obtained by dividing a count of matching bits, C_k, by a maximum number of matching bits, or length, N_k. *Glenn et al.* states that improved acoustic recognition is obtained by "normalizing" the scores to account for the consistency of the training patterns. (Column 6, Lines 66 to 68) It would have been obvious to one having ordinary skill in the art to normalize the score of similarity measures by dividing a similarity measure by a normalization score that varies in dependence upon the length of the stored sequence as taught by *Glenn et al.* in a speech recognition system and method of *Chou et al.* for a purpose of improving speech recognition by accounting for the consistency of training patterns.

Concerning claims 41 and 88, *Glenn et al.* discloses that the normalization scores, N_k, are generated by a maximum number of matching bits between the event encoding output and the candidate reference patterns stored in a reference pattern memory 60. (Column 6, Lines 31 to 65: Figure 1) The candidate reference patterns stored in reference pattern memory 60 are "a stored sequence of annotation labels".

Implicitly, each reference pattern has its own characteristic length, so that, in general, a normalization score, N_k, varies for each reference pattern.

Allowable Subject Matter

6. Claims 6 to 19, 26 to 29, 42 to 45, 47, 53 to 66, 74 to 76, 89, 91 to 92, and 94 to 95 are objected to as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims.

Response to Arguments

7. Applicants' arguments filed 28 December 2005 have been considered but are moot in view of the new grounds of rejection.

Conclusion

8. The prior art made of record and not relied upon is considered pertinent to Applicants' disclosure.

Goldberg ('261) and Brown et al. disclose related art.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Martin Lerner whose telephone number is (571) 272-7608. The examiner can normally be reached on 8:30 AM to 6:00 PM Monday to Thursday.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, David R. Hudspeth can be reached on (571) 272-7843. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

ML
3/19/07

Martin Lerner
Martin Lerner
Examiner
Group Art Unit 2626